

SOCIAL NETWORKS AND ETHNIC NICHE: AN ECONOMETRIC ANALYSIS OF THE MANUFACTURING SECTOR IN SOUTH AFRICA

ANDRE HOFMEYR*

Abstract

This paper analyses the link between social networks and ethnic occupational niches in the manufacturing sector in South Africa. To this end, it employs the methodology of Bertrand *et al.* to minimise the omitted variable bias induced by standard approaches investigating network effects and adopts Model's concentration index to define an ethnic niche. The results indicate that 25% of the sample is employed in ethnic niches in the manufacturing sector, but that niche employment varies markedly by language group. In addition, certain language groups tend to be clustered in advantageous niches where monthly income and skill levels are relatively high, while others occupy disadvantageous niches where monthly income and skill levels are relatively low. A number of different econometric specifications find strong evidence of social network effects. This highlights the role that these networks play in forming ethnic niches in the manufacturing sector in South Africa.

JEL Classification: J62, Z13

Keywords: Social network, ethnic niche, South Africa

1. INTRODUCTION

Social networks matter for labour market outcomes,¹ but to what extent do they channel workers into jobs that become ethnic niches? This paper adopts an econometric approach, which mitigates a number of biases prevalent in other studies, to analyse the impact of social networks on ethnic occupational niches in South Africa.

An ethnic occupational niche is the concentration and specialisation of members of an ethnic group in a particular occupational activity. According to theory, these niches arise because of the members' ability to supply labour through social networks and due to the special skills, experiences and other attributes they possess that employers consider relevant when hiring job applicants (see Waldinger, 1996a,b; Wilson, 2000; Elliott, 2001). Although a large literature documents the existence of ethnic niches and the importance of social networks in channelling individuals into these occupations, few studies have investigated the impact of social networks on the probability of niche employment using large sector-wide datasets.

* School of Economics, Rm 5.59, Leslie Social Science Building, University of Cape Town, Cape Town, South Africa. E-mail: andre.hofmeyr@gmail.com

Financial support for the working paper version of this article is gratefully acknowledged from Economic Research Southern Africa (ERSA). The author would like to thank two anonymous referees for all of their comments and suggestions during the revise and resubmit process.

¹ For a useful review of the literature on social networks and labour market outcomes, consult Ioannides and Loury (2004).

Social network analyses typically rely on detailed information about individuals' contacts collected through surveys. However, given the limited scope of this data and the complete absence of it in many cases, scholars have devised strategies to proxy for social networks, given the labour market data at their disposal. The most common method defines an individual's social network by the neighbourhood he inhabits, and numerous studies have shown that individual outcomes are indeed correlated with the outcomes of the individual's neighbours.² Another approach uses the ethnic group of the individual to proxy for this social network (see Borjas, 1992, 1995). Analyses using this framework have also found correlations between the outcomes of the ethnic group and the outcomes of the individuals that comprise it. While suggestive of network effects, the results of these studies should be treated with caution, because they are likely influenced by unobserved characteristics of individuals, neighbourhoods and ethnic groups.

To minimise the bias induced by these strategies, this paper adopts the approach of Bertrand *et al.* (2000) to define an individual's social network. Given that individuals tend to have homophilous social ties, we use language group to proxy for the links between individuals within a neighbourhood.³ This defines the "quantity" dimension of our measure of social networks: individuals who live in an area with a high proportion of other individuals who speak their language will have a larger pool of potential contacts who they can rely on to find employment. However, contacts that are actually employed in a particular occupation are likely to exert a larger influence on an individual's probability of finding employment in that occupation. Thus, the number of potential contacts that an individual has who are employed in an ethnic niche in his area provides a measure of network "quality". We investigate whether the "quantity" and "quality" of contacts that an individual has influences his probability of securing employment in an ethnic niche.

This paper uses the 10% sample of the 2001 Census survey conducted by Statistics South Africa. We use a two-part strategy to define occupations in specific areas as niche or non-niche. We then use a linear probability model, a probit and a logit to investigate whether an individual's social network influences the likelihood of niche employment. Our estimation strategy controls for many of the common omitted variable biases that have plagued previous studies. Specifically, we include language group, area and

² Consult Jencks and Mayer (1990) for a review of the literature.

³ We use language as our proxy, rather than other measures like race, gender or religion, because we think it plausible that job-related information flows more quickly and effectively through individuals who speak a common language. Bakalian (1993) shows that friendship groups tend to sort across ethno-linguistic lines, and Alba (1990) asserts that one's home language is a crucial determinant of ethnic identity. Furthermore, we think it reasonable that individuals typically spend more time with, and thus acquire information from, individuals who live in the same locality and speak the same language. In South Africa, markers like race, religion or gender are too broad to accurately define an individual's social network because of the differences that exist between individuals within these categories. For example, a black South African who speaks IsiZulu does not necessarily understand or associate with a black South African who speaks IsiXhosa even if they live in the same area. Finally, Orman (2008) argues that throughout South Africa's history, language planning has been a core feature of successive governments' attempts to construct and manipulate group identities.

occupation fixed effects in our estimation frameworks to control for these confounds. Our results suggest that social networks have a large and significant effect on the probability of niche employment.

The paper is structured as follows. Section 2 reviews the literature on social networks, labour market outcomes and ethnic niches. Section 3 discusses the empirical methodology adopted, and section 4 describes the data. Section 5 presents the empirical results, outlining the large role that social networks play in creating and maintaining ethnic niches. Section 6 conducts specification and robustness checks and section 7 concludes.

2. SOCIAL NETWORKS AND ETHNIC NICHES: A REVIEW OF THE LITERATURE

2.1 *Social Networks*

The importance of social networks for labour market outcomes has been increasingly recognised by the economics discipline. Social networks matter because they facilitate the transmission of job-related information between individuals. Unlike standard approaches to modelling job-search behaviour, social network models incorporate the information spillovers and complex interactions between individuals that are prevalent in contemporary labour markets.

An empirical regularity motivating this change in emphasis, which has been identified in both the economic and sociological literature from as early as 1960, is that on average 50-60% of workers obtain their jobs through personal contacts (Rees, 1966; Holzer, 1987; Staiger, 1990; Montgomery, 1991; Granovetter, 1995; Topa, 2001). Furthermore, 40-50% of employers use their employees' social contacts to fill job openings (Holzer, 1987). Informal recruitment methods have also been identified as improving the employer–employee match: individuals recruited through personal contacts are less likely to quit (Datcher, 1983; Devine and Kiefer, 1991) and have longer tenure on these jobs (Simon and Warner, 1992).

Theoretically, social networks matter for employment outcomes because of the functions that personal contacts play in labour markets (Elliott, 2001). These functions are twofold. Personal contacts can provide timely information about employment opportunities that are not widely or publicly known. Furthermore, contacts can pass on information to employers about a potential employee that increases his likelihood of being hired.

While the benefits to jobseekers from using social networks are quite obvious, the benefits to employers are less so. After all, if employers can fill job vacancies through formal means then what advantages do social networks confer? Fernandez *et al.* (2000) list five mechanisms that make hiring through social networks attractive. First, employers can enlarge the pool of applicants by drawing on the referrals of incumbents. Second, owing to the homophilous nature of incumbents' ties (Ullman, 1966; Rees and Schultz, 1970; Granovetter, 1995), the employer who has already screened the quality of his employees will be more disposed to hiring referrals because they are likely to be of a similar quality. In this sense, incumbents' referrals reduce the problems associated with information asymmetries. Third, employees – for fear of damaging their reputations – will only refer qualified candidates, which reduces the costs of screening applicants. Fernandez *et al.* (2000) refer to these three mechanisms collectively as the “richer pool” argument: insider referrals provide a larger and better pool of applicants. The fourth mechanism that makes

hiring through informal means beneficial to the employer, is that referrals are likely to provide a better match between employer and employee. Not only do incumbents pass on valuable information to employers about the individuals that they have referred, which is often difficult to obtain through formal recruitment procedures, but they also provide tacit information to candidates about the job for which they are applying. Finally, the benefits of referral hiring are present even after the job has been filled. The idea is that incumbents who referred an applicant will help the newcomer get acquainted with the organisation and even provide him with some on-the-job training, thereby boosting his productivity. In sum, there are numerous benefits to employers and employees from using social networks in the hiring process.

While a large empirical literature – from which the stylised facts presented above were drawn – regarding social networks and labour market outcomes exists, analyses of the impact of social networks on job search and employment in South Africa are somewhat limited. Most of the work that has been done defines the social network as the number of household members that are employed (see Wittenberg and Pearce, 1996; Mlatsheni and Rospabe, 1999; Schoer and Leibbrandt, 2006). Wittenberg and Pearce (1996) find that this network positively affects individuals' access to jobs, while Mlatsheni and Rospabe (1999) find that it increases the probability of youth being in wage employment. These studies, however, suffer from omitted variable bias, which makes their estimates unreliable, a point recognised by Burns *et al.* (2009). For their analysis, they instead adopt the approach of Bertrand *et al.* (2000) and define the social network in terms of age, language group and geographical proximity. Their estimation strategy also includes geographic area and age–language cohort fixed effects to minimise the potential for omitted variable bias. Their findings suggest that social networks may increase employment probabilities by an estimated 3-12%.

The preceding discussion highlights the importance of social networks for labour market outcomes. Given the widespread unemployment and stark income inequality in South Africa, it is crucial that we understand the impact of these networks on individuals' employment probabilities. This paper follows the approach of Bertrand *et al.* (2000) to analyse the extent to which social networks in South Africa create ethnic niches in certain occupations. This is relevant to the labour market situation in this country insofar as ethnic niches promote or hinder upward social mobility.

2.2 *Ethnic Niches*

Up to this point, the term “ethnic occupational niche” has been used glibly to refer to co-ethnic concentrated workplaces. However, there is widespread confusion in the literature regarding distinctions between the terms “ethnic economy”, “ethnic enclave” and “ethnic niche”, which are artefacts of different disciplines.⁴ While there are important differences between these terms, the more general one, ethnic niche, will be used henceforth to describe an employment sector in which individuals from a specific ethnic

⁴ Ethnic niches are typically associated with the concentration and specialisation of members of an ethnic group in particular industrial/occupational activities (Wilson, 2000). Ethnic economy refers to the concentration of co-ethnic owners and workers in one or more related industries, the sum of which is an ethnic economy (Logan *et al.*, 1994). An ethnic enclave is essentially an ethnic economy, but one in which an ethnic group specialises in the production of particular goods, and there is spatial concentration (locational clustering) of ethnic enterprises (Wilson, 2000).

group are concentrated above a level one would expect based on their share of the total labour force of a local labour market⁵ (Wilson, 1997).

Ethnic niches are formed, at least theoretically, through social networks (Morales, 2004). As Waldinger (1996a) argues, to attain employment in an ethnic niche, informal social networks are particularly important. Given that individuals tend to have homophilous social ties,⁶ information regarding job opportunities is more likely to flow to individuals of the same ethnic group as incumbents. This creates an insider–outsider dynamic that becomes self-reproducing: as more individuals from a particular ethnic group are employed in an organisation, information about employment opportunities is passed on to ethnically defined insiders while outsiders are excluded. Put simply, ethnically segregated social networks lead to ethnically segregated workplaces. This process is characterised by path dependence and can lead to occupational closure⁷ (see Waldinger, 1996a).

Discrimination in the general labour market is often the precursor to the formation of niches in local labour markets because in the latter, discriminatory barriers are relatively low, or other ethnic groups are not overly represented (Granovetter, 1995; Sassen, 1995). This is closely linked to the notion of the job queue, first expounded by Thurow (1969, 1972, 1975). According to the theory, individuals in the labour market form part of a job queue, with the most highly qualified individuals at the top and the less qualified lower down. However, in ethnically diverse economies, like South Africa, groups of individuals are ranked according to employers' perceptions of the productivity of the group, with the human capital of an individual serving as another mediating factor. Thus, at each level of the skill hierarchy, members of the core cultural group receive preferential treatment, relegating other ethnic groups to the periphery of the labour market (Morales, 2004). Forced to accept whatever jobs are available once groups higher up in the queue have made their selections, marginalised ethnic groups form niches.⁸

Faced with discrimination in the general labour market, niches can provide a “protected environment” for co-ethnics to acquire skills and experience and receive equitable compensation (Waldinger, 1996a:95). However, niches can also trap ethnic groups in exploitative relationships, where their possibility for upward mobility is circumscribed (Bonacich, 1988). Granovetter (1973) argues that individuals with weak ties are more likely to receive information about employment opportunities from outside of their strong tie social ambit. By this view, ethnic niches – and in particular, ethnic

⁵ Refer to the discussion of the concentration index in section 3.2 of this paper.

⁶ Buhai and van der Leij (2006) refer to this phenomenon as an in-breeding bias.

⁷ Occupational closure refers to the situation where access to employment opportunities in a niche is restricted for individuals from other ethnic groups because the incumbent ethnic group has managed to establish informal regulatory mechanisms and procedures to protect the niche against encroachment from other groups. The idea is that employers and employees in a niche come to an, often tacit, agreement about hiring practices and promotional rules because of the benefits to both parties from using insider referrals (discussed above). Consequently, individuals from other ethnic groups typically face more exacting job entry requirements than co-ethnic insiders and internal promotion becomes the norm. As a result, the occupational niche effectively becomes closed to other ethnic groups.

⁸ While the formation of a niche is typically characterised by self-selection, niches also arise “spontaneously” when members of a particular ethnic group are concentrated in the “residual” jobs of the labour market (Wilson, 1997).

enclaves – limit individuals' ability to access information outside of their ethnic group, which may hinder their ability to secure jobs in the general labour market.⁹ Regardless of their long-term effects, once an ethnic group gains a foothold in a particular occupation, the process of selective recruitment, described above, then increases its representation in the occupation while excluding other groups.

Despite their tendency towards occupational closure, ethnic niches are not immutable. Once established, a niche associated with one ethnic group may become associated with another. Waldinger (1996a) identifies two scenarios through which this can occur. In the succession scenario, economic expansion increases the demand for ethnic groups further down the labour queue. As individuals from these groups move into the general labour market, ethnic groups still further down the queue may form niches in the previously restricted occupations. In the leapfrogging scenario, the quality of low-ranked ethnic groups may improve (*e.g.* educational attainment increases) which raises the demand for these groups and re-orders the labour queue, assigning other groups to the niches previously dominated by the now more prized groups. Whatever the mechanism, it is important to note that niches are not fixed entities and are prone to change over time.

The theory of ethnic niches is predominantly sociological and descriptive. Buhai and van der Leij (2006) are the first scholars to apply social network theory to dynamically model occupational segregation in the labour market. They construct a simple three-stage model of occupational segregation with two homogenous social groups applying for two different jobs. In the first stage, individuals strategically decide to acquire one of two specialised educations. In the second stage, individuals randomly form friendship ties, with the probability of forming a tie decreasing with social distance. In the final stage, individuals use their personal contacts to search for jobs. The model predicts complete occupational segregation when social networks are important for channelling workers into jobs and when these networks are formed assortatively.

There is now a growing body of literature, albeit primarily sociological, that analyses the formation and persistence of ethnic niches (see Lieberman, 1980; Morawska, 1990; Model, 1993; Logan *et al.*, 1994; Model and Ladipo, 1996; Waldinger, 1996a,b; Wilson, 1997, 2000; Elliott, 2001; Morales, 2004). For example, Lichter (2000) finds that the use of personal contacts in job search, channels co-ethnics into niches, which often become enclaves. Light and Gold (2000) find that once an ethnic niche has been established, incumbents exert enormous influence over the hiring of job applicants and typically direct employment to their co-ethnics. Although rich in qualitative information, most of these analyses are localised and specific to particular sectors or occupations in developed countries. This paper takes a more encompassing approach and uses econometric techniques to test for the presence of ethnic niches in the manufacturing sector of the South African labour market.

3. METHODOLOGY

This paper adopts the approach of Bertrand *et al.* (2000) to analyse the impact of social networks on the employment of individuals in ethnic niches and uses Model's (1993) concentration index to define an ethnic niche. Section 3.1 details the econometric

⁹ This is an important line of future research, but one that this paper does not address due to data limitations.

specification employed in this paper and discusses the construction of our network measure. Section 3.2 explains the calculation of the ethnic niche variable that we use.

3.1 Econometric Specification and Social Network Measure

Social networks affect individual behaviour primarily through two channels: information and norms (Bertrand *et al.*, 2000). With regards to ethnic niches, social networks can facilitate the transmission of job-related information to individuals of a specific ethnic group. In addition, the social norm channel may affect individuals' preferences and thereby increase the probability of individuals applying for jobs in which their ethnic group predominates. While these channels may differentially affect employment in ethnic niches, the data at our disposal prevents us from analysing them separately. Instead, we consider the extent to which an ethnic group forms niches and use this as a measure of their niche "culture" while acknowledging that this "culture" is shaped both by information and norms. The extent to which an ethnic group's behaviour influences the behaviour of an individual in that group is the social network effect that this paper investigates.

To explain how we measure the effect of social networks, assume that the true model governing employment in an ethnic niche is the following:

$$\Pr(\text{Empl}_{ijk}) = \text{Netw}_{ijk}\alpha^* + X_i^*\beta^* + Y_l^*\gamma^* + Z_j^*\delta^* + W_k^*\phi^* + \varepsilon_{ijk} \quad (1)$$

where i indexes individuals, l indexes language groups, j indexes areas and k indexes occupations. Empl_{ijk} is a dummy indicating employment in niche k ,¹⁰ Netw_{ijk} measures the information and social norms of the individual's contacts, X_i^* are observed and unobserved personal characteristics, Y_l^* are observed and unobserved language group characteristics, Z_j^* are observed and unobserved local area characteristics, W_k^* are observed and unobserved occupation characteristics, and ε_{ijk} is an error term.

The difficulty in estimating this specification is that data for the variable Netw_{ijk} is seldom collected. Ideally, one would want data on individuals' actual contacts and the extent of their social networks. In the absence of such data, scholars have typically used the mean characteristics of an individual's locality as a proxy for their social networks. In doing so, there is an implicit assumption that individuals are randomly distributed within the neighbourhood. According to this framework, one would estimate:

$$\Pr(\text{Empl}_{ijk}) = \overline{\text{Empl}}_{jk}\alpha + X_i\beta + \varepsilon_{ijk} \quad (2)$$

where $\overline{\text{Empl}}_{jk}$ represents mean neighbourhood employment in occupation k and X_i are observed individual characteristics. Although a large body of empirical research validates

¹⁰ Specifically, Empl_{ijk} equals one if individual i from language group l , living in area j , is employed in occupation k where occupation k is an ethnic niche dominated by i 's language group. Empl_{ijk} equals zero if occupation k is not dominated by the language group of individual i . It should be noted that this measure is restricted to employed individuals. Thus, Empl_{ijk} draws a distinction between those individuals employed in niches and those not employed in niches without reference to unemployed individuals. Consult section 3.2 for more information on the construction of our niche variable.

the notion that individual outcomes are strongly correlated with mean neighbourhood characteristics (see Jencks and Mayer, 1990, for a review of the literature), the problem with this approach is that it suffers from Manski's (1993) "reflection problem".

Manski (1993) identifies three channels through which individual behaviour is affected: (i) *endogenous interactions*, where the behaviour of an individual is affected by the behaviour of the group; (ii) *contextual interactions*, where an individual's behaviour is affected by exogenous characteristics of the group; and (iii) *correlated effects*, where individuals in the same group behave in a similar manner because they share similar characteristics or face similar institutional environments. The "reflection problem" arises because data typically does not allow one to readily distinguish between these effects, making causal inference misguided at best, vacuous at worst. The difficulty is that mean behaviour in a group is affected by the behaviour of the individuals comprising the group. Thus one cannot determine whether group behaviour actually affects individual behaviour or whether group behaviour is simply the aggregation of each individual's behaviour.

The "reflection problem" can be viewed as the outcome of three related omitted variable biases (Bertrand *et al.*, 2000). First, omitted personal characteristics may be associated with \overline{Empl}_{jk} . For example, individuals that live in areas with widespread unemployment may be less ambitious. Second, omitted neighbourhood characteristics may be correlated with \overline{Empl}_{jk} . For example, areas with a training centre for occupation k may raise an individual's probability of being employed in this occupation and therefore raise the mean employment in occupation k in the area. Furthermore, the specification suffers from a simultaneity problem in that any policy or shock that affects mean neighbourhood employment will lead to a positive estimate of α , regardless of whether social networks are in operation. Finally, omitted occupational characteristics may be associated with \overline{Empl}_{jk} . For example, differences in skill requirements for particular occupations may make certain neighbourhoods (say, ones with higher average skill levels) more likely to dominate the employment of these occupations, regardless of whether social networks in fact facilitate acquisition of these jobs. Furthermore, industrial agglomeration may make the supply of certain occupations greater in particular areas as compared with others. As these biases are all likely positive, a positive estimate of α does not necessarily imply the existence of social networks.

Another approach, pioneered by Borjas (1992, 1995), uses ethnic groups – rather than geographical proximity – to proxy for social networks. As individuals tend to have homophilous social ties, one would expect ethnicity to be an important determinant of social networks. In addition, this approach focuses on the effect of the previous generation's outcomes on the current generation's outcomes. Therefore, the mean outcomes of the ethnic group in the previous generation is used to construct $Netw_{ijk}$. In the context of ethnic niches, one could estimate the following equation:

$$\Pr(Empl_{ijk}) = \overline{Empl}_{(-1)lk} \alpha + X_i \beta + Y_l \gamma + \varepsilon_{ijk} \quad (3)$$

where $\overline{Empl}_{(-1)lk}$ is the mean employment of ethnic group l in occupation k in the previous generation, and Y_l are observed language group characteristics.

However, this approach also suffers from three omitted variable biases. First, omitted personal characteristics may be associated with $\overline{Empl}_{(-1)lk}$. Second, omitted ethnic group characteristics may be correlated with $\overline{Empl}_{(-1)lk}$. For example, high levels of

discrimination may preclude certain language groups from obtaining employment in particular occupations, thereby concentrating their employment in others. This would show up as a positive estimate of α without directly capturing the effect of social networks. Finally, omitted occupational characteristics may be correlated with $\overline{Empl}_{(-i)lk}$.

This paper exploits both geographic and ethnic variation to construct a measure of social networks. Specifically, we use an individual’s language group and magisterial district to proxy for his social network. In addition, we include fixed effects for language groups, magisterial districts and occupations in our estimation frameworks to minimise the problems of omitted neighbourhood, ethnic group and occupational characteristics that have plagued previous studies.

We construct $Netw_{ijk}$ using the number of people an individual interacts with as well as the knowledge and attitudes of those people with respect to ethnic niches. Our measure therefore includes both a “quantity” and “quality” dimension of an individual’s potential contacts. Assuming individuals primarily interact with individuals of their language group, we therefore define:

$$Netw_{ijk} \approx \left(\begin{matrix} \text{density of language} \\ \text{group } l \text{ in area } j \end{matrix} \right)_{ij} \times \left(\begin{matrix} \text{ethnic niche knowledge} \\ \text{and attitudes of others} \\ \text{from language group} \\ l \text{ who live in area } j \end{matrix} \right)_{ijk}$$

The density of language group l in area j is a measure of contact availability, denoted by CA_{lj} .¹¹ CA_{lj} is thus our “quantity” measure. The second term in the construction of our network measure suggests that we should proxy for the “quality” of an individual’s potential contacts with the mean employment in area j of language group l in occupation k (excluding individual i), which we refer to as $\overline{Empl}_{(-i)ljk}$. However, using $\overline{Empl}_{(-i)ljk}$ can introduce another source of omitted variable bias, because it may reflect unobserved characteristics that an individual shares with members of his language group living in

¹¹ Specifically, CA_{lj} is the proportion of individuals from language group l in area j divided by the proportion of individuals from that language group in South Africa. The contact availability

measure is therefore defined as: $\frac{C_{lj}/A_j}{N_l/T}$, where C_{lj} is the number of people from language group

l in area j , A_j is the total number of people in area j , N_l is the total number of people who belong to language group l and T is the total number of people in the country.

Bertrand *et al.* (2000) typically use the log of this ratio in their calculations. Their rationale for doing so is that this prevents small magisterial districts from being underweighted in the regressions. However, as the CA measure is constructed by dividing the ratio of individuals of a particular language in a particular area by the ratio of the total number of individuals in this language group in the country, the measure is scale free. Thus, we use the unlogged version of CA, but test the robustness of our results to different choices of this measure in section 6.

Note that this measure refers to potential contacts and not actual contacts because the data lacks social network information. It should also be noted that this measure includes both employed and unemployed individuals, because one’s contact availability is defined simply by the “quantity” of potential contacts. The other term in our network measure captures the “quality” dimension and thus only includes employed individuals.

his area (Manski's *correlated effect*). Consequently, we replace $\overline{Empl}_{(-i)jk}$ with the mean employment of language group l in occupation k , \overline{Empl}_{lk} .¹²

We therefore estimate:

$$\Pr(Empl_{ijk}) = (CA_{ij} * \overline{Empl}_{lk})\alpha + X_i\beta + \gamma_l + \delta_j + \phi_k + CA_{ij}\theta + \varepsilon_{ijk} \quad (4)$$

where γ_l , δ_j and ϕ_k are fixed effects for language groups, magisterial districts and occupations, respectively. Note that the interaction of the "quantity" of potential contacts CA_{ij} and the "quality" of potential contacts \overline{Empl}_{lk} is used to proxy for an individual's social network: they provide a measure of how social networks influence individual behaviour. A positive and significant estimate of α provides evidence of network effects. We also include CA_{ij} as a control variable, but omit \overline{Empl}_{lk} because the language group fixed effects γ_l and occupation fixed effects ϕ_k incorporate it.¹³

This approach allows us to control for many of the common omitted variable biases prevalent in studies of social networks. First, magisterial district fixed effects capture differences in local areas, such as the extent of ethnic niching. Second, language group fixed effects capture omitted differences in language groups, such as the extent of discrimination they face. Third, occupation fixed effects capture omitted differences in occupations, such as varying skill requirements. Finally, by including CA_{ij} as a covariate, we control for omitted personal characteristics that may be associated with CA_{ij} . For example, an unobserved personal characteristic such as ambition may reduce the likelihood of an individual being employed in an ethnic niche, as well as living among his own language group.

Despite these controls, another potential source of omitted variable bias is still present: omitted personal characteristics that are correlated with the network term $CA_{ij} * \overline{Empl}_{lk}$. Including CA_{ij} captures fixed differences between individuals that choose to live among their own language group and those that do not. However, these differences may vary by language group: individuals in a specific language group may differentially self-select away from their own language group as compared with individuals in other language groups. For example, living away from your language group may indicate that you have managed to break into the general labour market if you are from a language group that is heavily concentrated in ethnic niches. Alternatively, selecting away from your language group, if it is one that is not disproportionately represented in ethnic niches, may signal the reverse. This would affect the estimate of α , thereby biasing the results. To control for this

¹² Note that the mean employment of language group l in occupation k is taken in deviation from the mean employment of all language groups in occupation k . Doing so removes what the language group fixed effect is capturing. Our measure is therefore $\overline{Empl}_{lk} - \overline{Empl}_k$. In results not reported here, which are available upon request, we also used the mean employment of language group l in occupation k not in area j $\overline{Empl}_{l(-j)k}$ as a measure of contact quality. Given the large sample used in the analysis, the results do not change significantly.

¹³ Technically, the interaction of γ_l and ϕ_k would capture the effect of mean employment of language group l in occupation k . This interaction is not included for the sake of parsimony of our specification. Furthermore, the interaction term would likely introduce significant collinearity in our model.

problem, we include a variable that indicates whether a person relocated between the 1996 and 2001 Census. Doing so should minimise the bias induced from individuals' differential self-selection away from their language groups.¹⁴

3.2 Ethnic Niche Measure

The preceding discussion has focussed exclusively on the right hand side of our econometric specification. Clearly, we now need to explain how we measure the dependent variable used in the analysis. The dependent variable ($Empl_{l,j,k}$), which we refer to as "employment in a niche", is dichotomous and equals one if individual i from language group l , living in magisterial district j is employed in occupation k , where occupation k is dominated by the language group of individual i in area j . This implies that the individual is employed in his language group's niche. It equals zero if occupation k is dominated by a language group other than individual i 's language group. In other words, "employment in a niche" equals one only if individual i 's language group is over-represented in occupation k in area j . To bring clarity to the terms "dominated" and "over-represented", we define an ethnic niche using a concentration index:

$$CI_{l,j,k} = \frac{e_{ljk} / o_{ljk}}{e_{lj} / o_{lj}}$$

where $CI_{l,j,k}$ is the concentration index for language group l living in area j and employed in occupation k ; e_{ljk} is the total number of individuals from language group l living in area j employed in occupation k ; e_{lj} is the total number of individuals of language group l living in area j ; o_{ljk} is the total number of individuals from other language groups living in area j and employed in occupation k ; and o_{lj} is the total number of individuals from other language groups living in area j (Wilson, 2000).

Model (1993) originally developed this approach for classifying occupations as niches and argued that an ethnic niche exists if the concentration index for language group l living in area j employed in occupation k is greater than 1.5. This approach was later adopted by Waldinger (1996a) and Wilson (2000), with the latter scholar adding the criterion that at least 20 workers be employed in this occupation in the sample for it to constitute a niche. The concentration index is an odds ratio, which implies that for a value of 1.5, language group l is 1.5 times more likely to be concentrated in occupation k than all other language groups in the area. As Wilson (2000) notes, the choice of 1.5 is arbitrary, but it sets a lower limit for the extent to which an ethnic group is concentrated in an occupation.

We adopt a two-part strategy for defining an ethnic niche. First, if the concentration index for at least 20 workers in language group l in area j employed in occupation k is greater than 1.5, we provisionally define the language group as occupying a niche. Second, we compare the concentration indices of all language groups whose values exceed 1.5 in area j and occupation k , and define an ethnic niche according to which group has

¹⁴ Inclusion of this variable requires some caveats. Note that an implicit assumption is made that if individuals relocate, then they sever their network. While this is true, at least in the short run, they may relocate to a nearby area where they interact with their language group, implying that they are still moving in roughly the same social networks. Unfortunately, there is no way to control for this given the data at our disposal. We thank an anonymous referee for pointing this out.

the highest concentration index.¹⁵ This ensures that the language group that we classify as occupying a niche unambiguously dominates the employment of a particular occupation because it is more concentrated than all other language groups. Only once a language group living in an area and employed in an occupation passes the two-part test is the occupation classified as an ethnic niche and the dependent variable assigned a value of 1

4. DATA

This paper uses the 10% sample of the 2001 Census survey conducted by Statistics South Africa. Since our focus is on the working age population, we exclude all individuals younger than 15 and older than 65. In addition, we exclude all non-housing units.¹⁶ We use magisterial district¹⁷ as our geographic indicator and define ethnicity by language group.¹⁸ The 2001 Census uses the South Africa Standard Classification of Occupations (SASCO) down to the three-digit level. We classify occupations at the two-digit level¹⁹ to allow for enough occupational heterogeneity in the sample and to ensure that our niche measures are not meaningless.²⁰

¹⁵ There were 39 cases where two language groups of 20 individuals or more had concentration indices above 1.5. As discussed above, the language group with the highest concentration index was defined as occupying the ethnic niche in this occupation.

¹⁶ Specifically, all prisons, residential hotels, student residences, homes for the aged, workers' hostels, tourist hotel/motels, hospital/medical facility/clinics, childcare institutions, homes for the disabled, boarding school hostels, initiation schools, convents/monasteries/religious retreats, defence force barracks, prisons/correctional institutions, community or church halls, refugee camps, and homeless shelters are excluded. This is done so that measures that rely on household size are not biased by the often huge number of "household" members in these non-housing units. For example, the Census records the number of household members of a prisoner as the number of individuals in the prison. In addition, some of the individuals in these non-housing units do not form part of the labour market, those in prison or mental health facilities for example.

¹⁷ There are 354 magisterial districts in the sample, which vary markedly in size from 85 to 35,334 individuals. A lower level of disaggregation was precluded by lack of information.

¹⁸ There are 11 official languages in South Africa. Data on all other languages spoken in the country is lumped into an "other" category, which is excluded from the analysis because more detailed information would be required to proxy for these language group's social networks.

¹⁹ To illustrate the difference between the levels at which occupations are classified, the following example is presented. A one-digit classification of an occupation is "Elementary occupations". The two-digit level breaks this classification down further into "Sales and services elementary occupations", "Agricultural, fishery and related labourers", and "Mining, construction, manufacturing and transport labourers". The three-digit level breaks each of these categories down further. For example, the two-digit level classification of "Agricultural, fishery and related labourers" at the three-digit level is "Agricultural, fishery and related labourers NFD", "Agricultural, fishery and related labourers", and "Agricultural, fishery and related labourers NEC".

²⁰ Elaboration on this point is clearly required. If we were to classify occupations at the one digit level, this would provide too broad a measure to classify occupations as ethnic niches, because individuals are matched to jobs, not broad occupational categories (Elliott, 2001). As discussed later, we limit the sample to the manufacturing sector. Consequently, classifying occupations at the three-digit level would be vacuous because in numerous areas, only one individual from one language group is employed in a particular occupation. This would make the concentration index for those individuals approach infinity, thereby defining a niche as the employment of one individual from one language group in one area.

Table 1. Summary statistics

Variable	Mean	Standard deviation
Employment in a niche	0.25	0.43
Contact availability	3.90	5.04
Black	0.56	0.50
Coloured	0.17	0.38
Asian/Indian	0.08	0.27
White	0.19	0.39
Male	0.66	0.47
Rural	0.14	0.35
Relocated between 1996 and 2001 Census	0.19	0.39
Access to a telephone	0.98	0.14
Disability	0.02	0.14
Married	0.64	0.48
Age	37.43	10.34
Years of education	9.67	4.06
Fraction of other adult household members employed in manufacturing	0.08	0.16

As discussed previously, the size of social networks is measured by contact availability CA_{ij} . We use the full sample to construct our measure of contact availability, following Bertrand *et al.* (2000), but limit the sample to the manufacturing sector for our regressions.²¹ This is done for two reasons. First, the manufacturing sector is the second largest contributor to South Africa's GDP (Statistics South Africa, 2006), and it employs 13% of South Africa's workforce. It is therefore broadly representative of the South African economy. Second, focussing on the manufacturing sector allows us to analyse ethnic occupational niches at a level of disaggregation that would be infeasible using the full sample due to the computational demands involved.

Table 1 presents summary statistics for the main variables used in the analysis. Of particular interest is that 25% of our sample is employed in ethnic niches in the manufacturing sector. Although quite high, this value falls well within the range of other studies (see Waldinger, 1996a; Wilson, 2000; Elliott, 2001).

As is evident, the manufacturing sector is dominated by males in their late thirties and is highly concentrated in urban areas. We find that black South Africans are under-represented in manufacturing relative to their share of the total population, whereas white, coloured and Asian/Indian South Africans are over-represented in this sector.²² This is clearly a legacy of Apartheid-era employment and education policies that gave preferential treatment to the population groups that are over-represented in manufacturing.

²¹ In doing so, we implicitly assume that individuals from one's language group living in one's area can provide information on employment prospects in the manufacturing sector even if they are not employed in this sector. This assumption is reasonable if social networks facilitate the transmission of job related information outside the confines of an individual's employment sector. For example, assume that person X does not know person Y. However, the father of person X (who is not employed in manufacturing) is told by person Y (who is employed in manufacturing) that the company he works for is looking to hire individuals. As a result, person X becomes aware of a job opening even though his direct social contacts are not employed in manufacturing. In addition, if an individual's contacts are not employed in manufacturing, this conveys valuable information as well; specifically, that it may be difficult to get a job in manufacturing. We thank an anonymous referee for pointing this out.

²² Black individuals constitute 79.24%, coloured individuals 9.18%, Asian/Indian individuals 2.52% and white individuals 9.06% of South Africa's total population.

Table 2. Summary statistics by language group

Language	Group size in full sample	Group size in manufacturing	Employment in a niche	Age	Years of education	Fraction of other adult household members employed in manufacturing	Relocated between the 1996 and 2001 census
Afrikaans	317,891	21,538	0.23	36.27	10.60	0.10	0.19
English	200,689	20,146	0.43	38.97	11.90	0.11	0.22
IsiNdebele	36,656	1,035	0.00	37.91	8.11	0.05	0.18
IsiXhosa	386,592	8,683	0.21	37.32	8.49	0.05	0.19
IsiZulu	500,289	18,435	0.32	36.95	8.51	0.06	0.15
Sepedi	203,221	5,662	0.11	38.09	8.50	0.05	0.20
Sesotho	186,128	6,684	0.08	37.14	8.51	0.05	0.17
Setswana	191,673	5,560	0.13	38.41	8.75	0.06	0.17
Siswati	58,440	1,816	0.09	36.93	7.44	0.07	0.16
Tshivenda	49,080	1,138	0.07	37.50	8.05	0.05	0.18
Xitsonga	96,821	2,711	0.04	36.59	7.39	0.06	0.19

Table 2 presents selected summary statistics for the 11 language groups used in our analysis. We find that employment in a niche varies markedly by language group. At one extreme, 43% of English speakers and 32% of IsiZulu speakers are employed in occupational niches. At the other, only 4% of Xitsonga speakers and 0 IsiNdebele speakers are employed in niches.

Approximately one-fifth of the individuals in our sample relocated between the 1996 and 2001 census, but this too varies by language group. It is therefore important to use this variable in our models to control for the differential self-selection of individuals away from their language groups.

The table indicates that English and Afrikaans speakers dominate the employment of the manufacturing sector. Furthermore, these two language groups have the highest levels of education, on average, as well as the highest fraction of other adult household members employed in manufacturing.²³

Table 3 presents detailed information on the extent of ethnic niching of each language group in each occupation. In addition, it lists the number of individuals employed in each occupation and their associated monthly income.²⁴ Each cell in the table, other than those listing the number of observations and mean monthly income, represents the fraction of individuals from language group l employed in a niche in occupation k . Interestingly, the table reveals that no ethnic niches exist in seven of the occupational categories.²⁵ However,

²³ A decision had to be made as to whether we should construct the variable “fraction of other adult household members employed” as opposed to “fraction of other adult household members employed in manufacturing”. We chose the latter option to capture the direct effect on an individual’s likelihood of niche employment in the manufacturing sector from having other household members employed in manufacturing rather than the more indirect effect from simply having other household members employed, regardless of the sector in which they work.

²⁴ Please note that the Census asked individuals to select the gross income category into which they fell. To compute the mean income of individuals in a particular occupation, the midpoint of the income category was used. For the highest income category, which is unbounded, a conservative value of R250,000 was used.

²⁵ The seven occupations are: “Legislators and senior officials”, “Life science and health professionals”, “Teaching professionals”, “Life science and health associate professionals”, “Teaching associate professionals”, “Subsistence, agricultural and fishery workers”, and “Agricultural, fishery and related labourers”.

Table 3. The extent of ethnic niching by language group and occupation

Occupation	Observations	Mean income	Language										Occupation mean				
			Afrikaans	English	IsiNdebele	IsiXhosa	IsiZulu	Sepedi	Sesotho	Setswana	Siswati	Tshivenda		Xitsonga			
Legislators and senior officials	94	R15,470.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Corporate managers	3,290	R13,266.20	0.20	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49
General managers	2,392	R13,589.80	0.17	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44
Physical, mathematical and engineering science professionals	2,174	R11,913.80	0.27	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27
Life science and health professionals	158	R14,244.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Teaching professionals	91	R8,494.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other professionals	2,352	R15,260.20	0.07	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46
Natural and engineering science associate professionals	2,848	R5,936.59	0.20	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.19
Life science and health associate professionals	150	R7,973.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Teaching associate professionals	34	R3,747.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other associate professionals	3,093	R7,557.65	0.06	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41
Office clerks	9,054	R4,323.08	0.19	0.66	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30
Customer service clerks	1,219	R2,490.89	0.10	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
Personal and protective services workers	1,074	R2,519.55	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Models sales persons and demonstrators	3,284	R5,424.48	0.08	0.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28
Market-oriented skilled agricultural fishery workers	469	R2,369.72	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
Subsistence agricultural and fishery workers	2	R1,800.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Extraction and building trades workers	5,379	R2,886.26	0.05	0.04	0.00	0.22	0.00	0.10	0.19	0.19	0.15	0.10	0.00	0.00	0.00	0.09	0.10
Metal machinery and related trades	8,410	R3,204.69	0.26	0.07	0.00	0.00	0.00	0.06	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.12	0.10
Handicraft, printing and related trades workers	1,987	R3,184.70	0.00	0.32	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11
Other craft and related trades workers	6,704	R1,613.54	0.16	0.00	0.00	0.08	0.00	0.61	0.08	0.08	0.03	0.12	0.35	0.35	0.00	0.00	0.22
Stationary-plant and related operators	2,429	R2,466.28	0.19	0.00	0.00	0.20	0.19	0.28	0.00	0.28	0.10	0.00	0.00	0.00	0.00	0.00	0.14
Machine operators and assemblers	13,915	R2,012.58	0.68	0.01	0.00	0.22	0.00	0.57	0.06	0.08	0.08	0.19	0.00	0.00	0.00	0.00	0.36
Driver and mobile-plant operators	5,733	R2,158.03	0.09	0.00	0.00	0.43	0.44	0.44	0.42	0.06	0.25	0.25	0.00	0.00	0.00	0.13	0.26
Sales and services elementary occupations	2,680	R1,497.54	0.20	0.00	0.00	0.37	0.25	0.25	0.10	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.17
Agriculture, fishery and related labourers	449	R875.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mining, construction, manufacturing and transport labourers	11,459	R1,589.46	0.13	0.00	0.00	0.49	0.49	0.46	0.16	0.26	0.26	0.26	0.19	0.05	0.00	0.00	0.28
Other	2,485	R4,158.95	0.12	0.07	0.00	0.00	0.00	0.08	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.07
Group mean			0.23	0.43	0.00	0.21	0.32	0.32	0.11	0.08	0.08	0.13	0.09	0.07	0.04	0.04	0.25

the number of individuals employed in these occupations varies from 2 to 449, which negates the possibility of observing niches. In the other occupations, the extent of ethnic niching varies from a high of 49% (“Corporate managers”) to a low of 2% (“Personal and protective services workers”), with the fraction of individuals employed in a particular niche varying across language groups.

The table indicates that the three occupations with the highest fraction of individuals employed in ethnic niches (“Corporate managers”, “General managers” and “Other professionals”), are dominated by the employment of English, and, to a lesser extent, Afrikaans speakers. Specifically, 82% of English corporate managers, 75% of English general managers and 80% of English other professionals are employed in ethnic niches, whereas 20% of Afrikaans corporate managers, 17% of Afrikaans general managers and 7% of Afrikaans other professionals work in niches dominated by their language group. This is not surprising, given the employment preponderance of these language groups in manufacturing, their higher levels of education on average, and apartheid employment practices that reserved high-ranking positions for white South Africans.²⁶ Interestingly, these are highly skilled occupations, and the highest paying of those in which ethnic niches exist.²⁷ They can therefore be regarded as advantageous niches for the individuals that comprise them.

On the other hand, occupations like “Drivers and mobile-plant operators”, “Mining, construction, manufacturing and transport labourers”, and “Extraction and building trades workers” may be regarded as disadvantageous niches for the individuals that occupy them. These are relatively low-skilled occupations that are low paying,²⁸ and their niches tend to be concentrated among traditionally black language groups.²⁹ As is evident, IsiZulu and IsiXhosa speakers have the highest fraction of individuals employed in niches in these occupations.

Table 3 is particularly striking because it highlights the stark realities of South Africa’s labour market. The extent of ethnic niching varies markedly by language group and occupation. More importantly, perhaps, it shows that individuals form part of advantageous or disadvantageous niches. Traditionally white language groups, and particularly English speakers, tend to form part of advantageous niches where monthly income and the skills required for these jobs are relatively high. Traditionally black language groups, on the other hand, tend to be concentrated in disadvantageous niches where monthly income and skill requirements are relatively low. From the perspective of the labour market as a whole, if particular groups of individuals are trapped in ethnic niches where income and skills are low while others enjoy the protection of ethnic niches

²⁶ Of all white South Africans employed in manufacturing, 51.2% speak Afrikaans and 48.3% speak English. Although apartheid was based on a racial classification system, the high correlation between language and race necessarily implies that English and Afrikaans speakers were the primary beneficiaries of this policy.

²⁷ The mean income for corporate managers, general managers and other professionals is R13,266.20, R13,589.80 and R15,260.20, respectively.

²⁸ The mean income for drivers and mobile plant operators, mining, construction, manufacturing and transport labourers, and extraction and building trades workers is R2,158.03, R1,589.46 and R2,886.26, respectively.

²⁹ The language groups traditionally associated with black South Africans are: IsiXhosa, IsiZulu, IsiNdebele, Sepedi, Sesotho, Setswana, Siswati, Tshivenda and Xitsonga.

where income and skills are high, then income mobility may be severely limited. The fear is that these niches create an insider–outsider dynamic that becomes self-reproducing, something which may lead to occupational closure. This has implications for income inequality, human capital investments and affirmative action policies in South Africa.

For example, if ethnic niching patterns persist then individuals whose contacts are trapped in relatively low-skilled, low-paying occupations may not invest in human capital because this is not required in the occupation where their ethnic group predominates. As another example, ethnic niches may undermine affirmative action policies if certain language groups lack the networks necessary to obtain employment in particular occupations. It is with this background on the extent of ethnic niching in the manufacturing sector in South Africa that we now explore the role of social networks in channelling workers into jobs that become ethnic niches.

5. EMPIRICAL RESULTS

To estimate the specification in equation (4), we use a linear probability model (LPM), a probit and a logit.³⁰ Our model includes demographic controls, fixed effects for language groups, magisterial districts and occupations, a measure of contact availability CA_{ij} , and the interaction of CA_{ij} and the mean employment of the individual's language group in occupation k taken in deviation from the mean employment of all language groups in occupation k : $CA_{ij} * (\overline{Empl}_{ik} - \overline{Empl}_k)$. The demographic controls include three race dummies, age, age squared, years of education, years of education squared, a dummy indicating whether the individual lives in an urban or rural area, marital status, gender, a dummy for whether the individual relocated between the 1996 and 2001 census, the fraction of other adult household members employed in manufacturing, a dummy indicating whether the individual has a disability, and finally a dummy indicating whether the individual has access to a telephone.

Table 4 presents the estimates of our network coefficient as we include fixed effects for language groups, magisterial districts and occupations.³¹ We observe a monotonic decline in our network coefficient when fixed effects are added to the LPM specification. This is not the case for the logit and probit models, where it is only the introduction of occupation fixed effects that have a marked negative impact on the estimate of our network measure. Regardless of which specification you regard as a better model of the underlying data, this set of results highlights the importance of our estimation strategy, because without fixed effects, our network measure is clearly biased upward by omitted

³⁰ Please note that the logit and probit specifications were estimated on a subset of the manufacturing sample ($n = 77,791$). The rationale for doing this is the inclusion of so many fixed effects in the model, a number of which are perfect predictors of failure of the response probability ($n = 15,617$) and are thus dropped during estimation, which decrease the size of our sample. To make the estimates of our network measure comparable across the different fixed effect specifications, all observations that were perfect predictors of the response probability in our complete model were removed from the data so that the effect of introducing fixed effects is not confounded by changes in sample size.

³¹ Our models include all of the demographic controls listed above, as well as the non-interacted CA_{ij} measure, but their coefficient estimates are not reported in this table.

Table 4. Estimates of network coefficient as additional fixed effects are included

	Network measure (α)		
	LPM	Probit ϕ	Logit ϕ
No controls	0.544*** [0.007]	0.970*** [0.012]	1.404*** [0.017]
+Language group fixed effects	0.520*** [0.007]	1.018*** [0.012]	1.354*** [0.016]
+Magisterial district fixed effects	0.508*** [0.007]	0.998*** [0.012]	1.375*** [0.016]
+Occupation fixed effects	0.473*** [0.007]	0.849*** [0.011]	1.146*** [0.015]

Note: Standard errors in brackets.

LPM: linear probability model; ϕ : marginal effects are reported.

* significant at 10%; ** significant at 5%; *** significant at 1%.

occupation characteristics, and, to a lesser extent, omitted language group and magisterial district controls.

Table 5 presents our main results. The covariates generally display the expected signs, and, other than our network and contact availability measures, the magnitude of these estimates are similar across all specifications.³² Being female and living in a rural area increases the probability of being employed in a niche, presumably because males find it easier to obtain employment in the general labour market, and urban areas increase the range of employment possibilities for individuals. Interestingly, the mean income of females in niche employment (R3,387.96) is higher than the mean income of females in non-niche employment (R2,473.78) but the mean income of females in niche employment (R3,387.96) is far lower than the mean income of males in niche employment (R7,193.49). Non-parametric Mann–Whitney tests confirm that these differences are statistically significant (female non-niche – female niche: $z = -15.17$, $p = 0.00$; female niche – male niche: $z = -38.22$, $p = 0.00$). Thus, female niche employment has a beneficial impact on income when compared with female non-niche employment, but there is clearly a large gender gap in the income of males and females in niche employment.

A similar analysis can be conducted for individuals residing in rural areas. We find that niche employment in rural areas is associated with a lower income than non-niche employment in rural areas (R1,305.69 compared with R1,701.95) and niche employment in rural areas is associated with a far lower income than niche employment in urban areas (R1,305.69 compared with R6,117.17). Mann–Whitney tests confirm that these differences are statistically significant (rural non-niche – rural niche: $z = 7.87$, $p = 0.00$; urban niche – rural niche: $z = 44.33$, $p = 0.00$). Thus, although living in a rural area increases the likelihood of niche employment, these niches are disadvantageous, in terms of income, for the individuals that occupy them.

Being non-white lowers the probability of being employed in a niche. This finding is driven by the fact that white individuals are over-represented in ethnic niches given their

³² One expects the estimates on our network and contact availability measures to be larger in the probit and logit specifications because these are estimated on a subsample of the manufacturing sector data that excludes numerous observations that were perfect predictors of failure of the response probability.

Table 5. LPM, probit and logit estimates of the network coefficient

	Employment in a niche		
	LPM	Probit ϕ	Logit ϕ
Network measure	0.473*** [0.007]	0.849*** [0.011]	1.146*** [0.015]
Contact availability	0.013*** [0.001]	0.039*** [0.001]	0.032*** [0.001]
Rural dummy	0.025*** [0.005]	0.016*** [0.007]	0.014** [0.006]
Male	-0.029*** [0.004]	-0.034*** [0.005]	-0.033*** [0.005]
Individual is black	-0.062*** [0.011]	-0.074*** [0.013]	-0.063*** [0.013]
Individual is coloured	-0.059*** [0.005]	-0.056*** [0.006]	-0.045*** [0.006]
Individual is Asian/Indian	-0.083*** [0.007]	-0.069*** [0.007]	-0.050*** [0.007]
Individual relocated between 1996 and 2001 Census	-0.012*** [0.003]	-0.013*** [0.004]	-0.013*** [0.004]
Has access to telephone	-0.016* [0.009]	-0.013 [0.012]	-0.013 [0.012]
Has disability	-0.006 [0.008]	-0.006 [0.009]	-0.009 [0.009]
Married	-0.006 [0.004]	-0.006 [0.004]	-0.006 [0.005]
Married male	0.008 [0.005]	0.009 [0.006]	0.009 [0.006]
Age	0.001 [0.001]	0.000 [0.000]	0.000 [0.000]
Years of education	-0.001*** [0.000]	-0.017*** [0.005]	-0.015*** [0.005]
Age squared	0.000 [0.000]	0.000 [0.000]	0.000 [0.000]
Years of education squared	-0.000* [0.000]	0.002** [0.000]	0.002** [0.001]
Fraction of other adult household members employed in manufacturing	-0.011 [0.008]	-0.004 [0.009]	-0.002 [0.008]
Constant	0.207*** [0.044]		
Observations	93,408	77,791	77,791

Note: Standard errors in brackets.

LPM: linear probability model; ϕ : marginal effects are reported.

* significant at 10%; ** significant at 5%; *** significant at 1%.

share of the total population. The mean income of whites employed in ethnic niches (R14,676.78) is far higher than the mean income of non-whites employed in ethnic niches (R2,516.51). A Mann–Whitney z-statistic of -92.16 ($p = 0.00$) confirms that this difference is statistically significant. Thus, white South Africans not only benefit from easier to access to niche employment, but also from the income they earn in niche employment. Thus, to echo the results presented in the previous section, white South Africans who predominantly speak English and Afrikaans tend to be clustered in advantageous niches, whereas non-white South Africans tend to form part of disadvantageous niches.

If an individual relocated between the 1996 and 2001 census, this lowers the probability of niche employment. Clearly social networks take time to form, and therefore individuals that have recently relocated will find it more difficult to obtain niche employment. We find that higher years of education decrease the probability of niche employment. This squares well with the sociological literature, which suggests that less-skilled individuals are more likely to be employed in ethnic niches. Finally, the fraction of other household members employed in manufacturing lowers the probability

of niche employment. Although somewhat surprising, this finding supports our decision to construct the contact availability measure using data from the full census, rather than purely from the manufacturing sector. The negative estimate of this coefficient, although not significant, suggests that it is not only those directly employed in manufacturing that are relevant to the probability of your employment in a *particular niche* in the manufacturing sector.³³

As in Table 4, we find that our network measure is positive and significant, indicating the importance of social networks for niche employment. However, the estimates of our network measure differ markedly across the models. As discussed previously, one expects the value of the network coefficient to be higher in the logit and probit specifications because these are estimated on a smaller sample that excludes a number of observations that are perfect predictors of failure of the response probability. Nevertheless, the estimate of our network coefficient is positive and highly significant across all specifications. Thus social networks matter for niche employment in the manufacturing sector in South Africa.

6. SPECIFICATION AND ROBUSTNESS CHECKS

In this section, we test whether our results are sensitive to the specification of our network measure, as well as the sample used in our analysis. Table 6 presents the estimates of our network measure when we change its specification. As the table indicates, we find positive and significant network effects in all models.

Row (i) displays the estimates of our network measure from Table 5 for ease of comparison. In row (ii), we estimate our model by using the logged version of our contact availability measure: $\ln((C_{ij}/A_i)/(N_i/T))$. In row (iii), we use (C_{ij}/A_i) as our contact availability measure. Evidently, this specification does not weight the measure by the language group's share of the total population. Finally, in row (iv), we use (C_{ij}) as our

Table 6. Specification checks

		Network measure (α)		
		LPM	Probit ϕ	Logit ϕ
(i)	Network measure from Table 4	0.473*** [0.007]	0.849*** [0.011]	1.146*** [0.015]
(ii)	CA Measured in logs rather than levels	3.258*** [0.032]	3.795*** [0.040]	4.004*** [0.043]
(iii)	CA measured as (C_{ij}/A_i)	9.686*** [0.091]	10.801*** [0.118]	11.843*** [0.136]
(iv)	CA measured as (C_{ij})	0.0005*** [0.000]	0.0004*** [0.000]	0.0005*** [0.000]

Note: Standard errors in brackets.

LPM: linear probability model; ϕ : marginal effects are reported;

CA: contact availability.

* significant at 10%; ** significant at 5%; *** significant at 1%.

³³ An alternative explanation for this result is that households diversify with regards to employment. For example, if a number of family members are employed in different occupations in the manufacturing sector, then this increases the range of opportunities for individual i , thereby lowering his probability of niche employment. We are grateful to Justine Burns for suggesting this alternative interpretation to us.

Table 7. *Sample selection checks*

	Network measure (α)
	LPM
(i) Exclude English speakers	0.114*** [0.007]
(ii) Exclude English and IsiZulu speakers	0.109*** [0.007]
(iii) Exclude white South Africans	0.416*** [0.007]
(iv) Exclude women	0.427*** [0.008]
(v) Exclude rural areas	0.950*** [0.010]

Note: Standard errors in brackets.

LPM: linear probability model.

* significant at 10%; ** significant at 5%; *** significant at 1%.

contact availability measure to investigate whether there are changing returns to scale (Bertrand *et al.*, 1998). As the table shows, our results are robust to changes in the specification of our network measure across all three models.

Table 7 presents the estimates of our network measure when we change the sample used in the analysis. Unfortunately, the changes in sample size introduced numerical instability to the logit and probit optimisation processes, which prevented these models from converging. Consequently, we only report the results from the LPM specification. As discussed previously, English and IsiZulu speakers constitute the largest fraction of individuals employed in ethnic niches, raising concerns that our results are driven primarily by these two groups. In row (i) of the table, we exclude English speakers and continue to find positive and significant network effects although the magnitude of the effect is greatly diminished.³⁴ In row (ii), we also exclude IsiZulu speakers and witness a further decline in the coefficient on our network measure. Clearly, social networks are particularly important for these groups in obtaining niche employment. However, they also have a significant effect on the probability of niche employment for other groups. In row (iii), we limit the sample to non-whites and once again observe a decline in our network measure relative to its value estimated on the full manufacturing sample. Nevertheless, the coefficient is still positive and significant, indicating that social networks play an essential role for non-whites in securing niche employment even though the niches that these individuals occupy may be disadvantageous. As women are more likely to be find niche employment, see Table 5, we remove them from our sample to ensure that our results are not driven by this gender – results reported in row (iv). Again, we witness a decline in our network measure, but find that the coefficient is still large in magnitude and significant. Finally, in row (v), we limit the sample to urban areas. The network measure increases dramatically, indicating that social networks are particularly important in urban areas for securing niche employment. This is arguably because of the competitive nature of urban area labour markets which make social networks crucial for channelling workers into ethnic niches.

³⁴ Levinsohn (2007) finds that the return to speaking English in South Africa rose between 1993 and 2000. This result, coupled with the fact that English speakers have the highest fraction of individuals employed in ethnic niches, explains to some extent the large drop in the coefficient on our network measure when we exclude English speakers.

As Tables 6 and 7 make clear, our results are robust to changes in specification of the contact availability measure, as well as the sample used in the analysis.³⁵ This highlights the benefit of our estimation strategy, as well as the importance of social networks for the formation of ethnic niches.

7. CONCLUSION

This paper adopted the approach of Bertrand *et al.* (2000) to explore the impact of social networks on ethnic niche employment in the manufacturing sector in South Africa. The estimation strategy allowed us to control for numerous omitted variables that have plagued previous studies of social networks using non-network data. We defined ethnic niches using a two-part strategy and found that social networks have a large, positive and significant impact on the probability of niche employment across all of our econometric specifications. While networks are particularly important for certain language groups, they unambiguously increase the likelihood of niche employment for all language groups in all areas and in all occupations. The robustness of our results to changes in specification, as well as sample selection, bolsters our confidence that the findings are not the product of the vagaries of econometric estimation.

The results raise a number of important issues about South Africa's labour market. The finding that English and Afrikaans individuals tend to be clustered in advantageous niches, whereas traditionally black language groups tend to be concentrated in disadvantageous niches, highlights the segmented nature of the labour market. The fact that social networks are a key channel to niche employment is likely to reinforce this segmentation if human capital investments become endogenised, individuals fail to form social networks outside of their strong tie social ambit and policies that aim to redress labour market inequalities do not alter the contacts that individuals have.

With regards to affirmative action policies, Burger and Jafta (2006), using data from 1995 to 2004, find that the gap in occupational attainment between whites and non-whites has stabilised over the period under study, and that the racial wage gap between whites and blacks has narrowed at the 70th, 90th and 95th percentiles of the wage distribution. These results indicate that important changes are starting to take place in the labour market but that gains thus far have been limited. Although purely conjecture, one of the reasons why affirmative action has failed to have a large impact on occupational attainment and wages of non-whites is that these individuals lack the social networks that are helpful in securing employment in occupations dominated by English and Afrikaans individuals. The reinforcing effects that social networks have in forming and maintaining niches may explain why it has proved so difficult to alter long-run patterns of occupational segregation. Hopefully, a more integrated and equitable education system should help to enlarge individuals' social networks so that these niching patterns start to

³⁵ We also ran regressions to test whether the network effect differed according to the age of the individuals in our sample (results not reported). We constructed five age cohorts: less than 25, 26-35, 36-45, 46-55 and 56-65. Despite a marked increase in the coefficient on our network measure for individuals younger than 25 (the coefficient increased to 0.642), it remained fairly stable for the other age cohorts. The increase in our network measure for individuals younger than 25 indicates the importance of social networks in channelling young individuals into ethnic niches presumably because their credentials and lack of experience often make it difficult for them to obtain employment in the general labour market.

change. It is crucial to realise though that policies aimed at labour market redress must take into account the complex interactions and information spillovers that social networks engender. Clearly, this is not a prescription for government to intervene in the networks that individuals form, but rather a suggestion that the influence of networks be taken into account during policy formation.

Another interesting issue is whether recruitment by referral is an efficient employment strategy for the firms that use it. A number of studies, cited previously, have found that it is. However, in a recent paper, Fafchamps and Moradi (2009) analyse military records from Ghana and find that referred recruits were more likely to desert or be dismissed from the army than their unreferred counterparts. Thus, the referral system did not improve the unobserved quality of new recruits, which is the standard outcome in models of employee referral (see Montgomery 1991). If this is true of the South African labour market, then recruitment by referral may help individuals with the right networks to find employment, but may prevent other, potentially more productive, workers from doing so. Not only does this lower the productivity of firms, but it also limits the extent of integration in the workplace and the labour market as a whole. This is clearly an important line of future research but one that this paper does not address due to data limitations.

In sum, the manufacturing sector of the South African labour market is characterised by a high proportion of ethnic niches. These niches can keep some language groups trapped in low-skilled, low-paying occupations while they protect other language groups in high-skilled, high-paying jobs from occupational encroachment. Social networks are crucial to the formation and maintenance of these niches, and if public policies are to tackle occupational segregation, they must take into account the networks that are prevalent in the contemporary labour market.

REFERENCES

- ALBA, R. (1990). *Ethnic Identity: The Transformation of White America*. New Haven, CT: Yale University Press.
- BAKALIAN, A. (1993). *American-Americans: From Being to Feeling American*. New Brunswick, NJ: Transaction Publishers.
- BERTRAND, M., LUTTMER, E. and MULLAINATHAN, S. (1998). Network effects and welfare cultures. *NBER Working Paper Series*: 6832.
- . (2000). Network effects and welfare cultures. *Quarterly Journal of Economics*, 116(1): 351-377.
- BONACICH, E. (1988). The social costs of immigrant entrepreneurship. *Amerasia Journal*, 14: 119-128.
- BORJAS, G. (1992). Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics*, 107: 123-150.
- . (1995). Long-run convergence of ethnic skill differentials: The children and grandchildren of the great migration. *Industrial and Labour Relations Review*, XLVII: 553-573.
- BUHAI, S. and VAN DER LEIJ, M. (2006). *A Social Network Analysis of Occupational Segregation*. Timbergen Institute Discussion Paper.
- BURGER, R. and JAFTA, R. (2006). *Returns to Race: Labour Market Discrimination in Post-Apartheid South Africa*. Working Papers 04/2006, Stellenbosch University, Department of Economics.
- BURNS, J., GODLONTON, S. and KESWELL, M. (2009). Social networks, employment and worker discouragement: Evidence from South Africa. *Labour Economics*, doi: 10.1016/j.labeco.2009.08.007/.
- DATCHER, L. (1983). The impact of informal networks on quit behaviour. *Review of Economic Statistics*, 65(3): 491-495.
- DEVINE, T. and KIEFER, N. (1991). *Empirical Labor Economics: The Search Approach*. New York: Oxford University Press.
- ELLIOTT, J. R. (2001). Referral hiring and ethnically homogeneous jobs: How prevalent is the connection and for whom? *Social Science Research*, 30: 401-425.
- FAFCHAMPS, M. and MORADI, A. (2009). Referral and Job Performance: Evidence from the Ghana Colonial Army. BREAD Working Paper No. 238.
- FERNANDEZ, R., CASTILLA, E. and MOORE, P. (2000). Social capital at work: Networks and employment at a phone center. *American Journal of Sociology*, 105(5): 1288-1356.
- GRANOVETTER, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6): 1360-1380.

- . (1995). The economic sociology of firms and entrepreneurs. In A. Portes (ed.), *The Economic Sociology of Immigration*. New York: Russell Sage Foundation, 128-165.
- HOLZER, H. (1987). Informal job search and black youth unemployment. *American Economic Review*, 77(3): 446-452.
- IOANNIDES, Y. M. and LOURY, L. D. (2004). Job information networks, neighborhood effects and inequality. *Journal of Economic Literature*, XLII(December): 1056-1093.
- JENCKS, C. and MAYER, S. (1990). The social consequences of growing up in a poor neighbourhood. In L. Lynn and M. McGeary (eds), *Inner-City Poverty in the United States*. Washington, DC: National Academy Press, 111-186.
- LEVINSOHN, J. (2007). Globalization and the returns to speaking English in South Africa. In A. Harrison (ed.), *Globalization and Poverty*. Chicago, IL: University of Chicago Press, 629-646.
- LICHTER, M. (2000). Race, space, and the employment of less-educated immigrants and natives. *Dissertation Abstracts International, A: The Humanities and Social Sciences*, 60(7): 2683-A-2684-A.
- LIEBERSON, S. (1980). *A Piece of the Pie: Blacks and White Immigrants since 1880*. Berkeley, CA: University of California Press.
- LIGHT, I. and GOLD, S. (2000). *Ethnic Economies*. San Diego, CA: Academic Press.
- LOGAN, J., ALBA, R. and MCNULTY, T. (1994). Ethnic economies in metropolitan regions: Miami and beyond. *Social Forces*, 72: 691-724.
- MANSKI, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3): 531-542.
- MLATSHENI, C. and ROSPABE, S. (1999). Why is Youth Unemployment so High an Unequally Spread in South Africa. Unpublished mimeo.
- MODEL, S. (1993). The ethnic niche and the structure of opportunity: Immigrants and minorities in New York City. In M. Katz (ed.), *The Historical Origins of the Underclass*. Princeton, NJ: Princeton University Press, 485-510.
- and LADIPO, D. (1996). Context and opportunity: Minorities in London and New York. *Social Forces*, 75: 999-1024.
- MONTGOMERY, J. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review*, 81(5): 1408-1418.
- MORAWSKA, E. (1990). The sociology and historiography of immigration. In V. Yans-McLaughlin (ed.), *Immigration Reconsidered*. New York: Oxford University Press, 187-238.
- MORALES, M. (2004). Ethnic Niches, Pathway to Economic Incorporation or Exploitation? Labour Market Experiences of Latina/os. Unpublished mimeo.
- ORMAN, J. (2008). *Language Policy and Nation-Building in Post-Apartheid South Africa*. Vienna: Springer.
- REES, A. (1966). Information networks in labor markets. *American Economic Review*, 56: 559-566.
- and SCHULTZ, G. (1970). *Workers and Wages in an Urban Labor Market*. Chicago, IL: Chicago University Press.
- SASSEN, S. (1995). Immigration and local labor markets. In A. Protes (ed.), *The Economic Sociology of Immigration: Essays on Networks, Ethnicity, and Entrepreneurship*. New York: Russell Sage Foundation, 87-127.
- SCHOER, V. and LEIBBRANDT, M. (2006). Determinants of job search strategies: Evidence from the Khayelitsha/Mitchell's Plain survey. *South African Journal of Economics*, 74(4): 702-724.
- SIMON, C. and WARNER, J. (1992). Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings and tenure. *Journal of Labor Economics*, 10(3): 306-330.
- STAIGER, D. (1990). The Effect of Connections on the Wages and Mobility of Young Workers. Mimeo, MIT.
- STATISTICS SOUTH AFRICA. (2006). *Gross Domestic Product, Second Quarter: 2006*. Statistical release P0441.
- THUROW, L. (1969). *Poverty and Discrimination*. Washington, DC: Brookings Institution.
- . (1972). Education and economic equality. *Public Interest*, 28: 66-81.
- . (1975). *Generating Inequality: Mechanisms of Distribution in the U.S. Economy*. New York: Basic Books.
- TOPA, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies*, 68(2): 261-295.
- ULLMAN, J. (1966). Employee referrals: Prime tool for recruiting workers. *Personnel*, 43: 30-35.
- WALDINGER, R. (1996a). *Still the Promised City? African-Americans and New Immigrants in Post-Industrial New York*. Cambridge, MA: Harvard University.
- . (1996b). Ethnicity and opportunity in the plural city. In R. Waldinger and M. Bozorgmehr (eds), *Ethnic Los Angeles*. New York: Russell Sage Foundation, 445-471.
- WILSON, F. (1997). *Ethnic Concentrations and Labor Market Opportunities*. CDE Working Paper No. 97-17.
- . (2000). *Metropolitan Labor Markets and Ethnic Niching: Introduction to a Research Project*. Institute for Research on Poverty Discussion Paper No. 120400.
- WITTENBERG, M. and PEARCE, C. (1996). Youth unemployment: Some perspectives from the South African Living Standards and Development Survey. In L. Chrisolm, J. Favish, C. Harrison, P. Kgobe and S. Motala (eds), *Out-of-School Youth Report: Policy and Provision for Out-of-School and Out-of-Work Youth*. Johannesburg: Education Policy Unit, University of Witwatersrand, 101-130.